

# The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances

Ulrich, Werner; Nakadai, Ryosuke; Matthews, Thomas J.; Kubota, Yasuhiro

DOI:

[10.1016/j.ecocom.2018.07.002](https://doi.org/10.1016/j.ecocom.2018.07.002)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Ulrich, W, Nakadai, R, Matthews, TJ & Kubota, Y 2018, 'The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances', *Ecological Complexity*, vol. 36, pp. 110-116.

<https://doi.org/10.1016/j.ecocom.2018.07.002>

[Link to publication on Research at Birmingham portal](#)

## Publisher Rights Statement:

Checked for eligibility: 23/08/2018

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

**The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances**

Werner Ulrich<sup>1\*</sup>, Ryosuke Nakadai<sup>2</sup>, Thomas J. Matthews<sup>3,4,5</sup>, Yasuhiro Kubota<sup>6</sup>

<sup>1</sup>Department of Ecology and Biogeography, Nicolaus Copernicus University, Toruń, Poland,

<sup>2</sup>Faculty of Science, University of the Ryukyus, Nishihara, Japan

<sup>3</sup>School of Geography, Earth and Environmental Sciences, University of Birmingham, UK

<sup>4</sup>CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Univ. dos Açores – Depto de Ciências e Engenharia do Ambiente, PT-9700-042, Angra do Heroísmo, Açores, Portugal

<sup>5</sup>Birmingham Institute of Forest Research, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

<sup>6</sup>Marine and Terrestrial Field Ecology, Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Japan

E-mails: WU: [ulrichw@umk.pl](mailto:ulrichw@umk.pl); RN: [r.nakadai66@gmail.com](mailto:r.nakadai66@gmail.com); TM: [T.J.Matthews@bham.ac.uk](mailto:T.J.Matthews@bham.ac.uk); YK: [kubota.yasuhiro@gmail.com](mailto:kubota.yasuhiro@gmail.com)

\* Corresponding author

Running title: Weibull fits to species relative abundance

**Author contribution**

WU developed the theoretical background, analyzed the data, and wrote the first draft. YK and RN provided the forest data. TM contributed theoretical background. All authored contributed significantly to the final text version.

## Abstract

The study of species abundance distributions (SADs) needs a precise modelling of their drivers and ecological implications. We introduce the two-parameter Weibull distribution as a versatile tool to fit various kinds of observed SADs and to compare observed and theoretically expected values at the species level. We show that the shape and the scale parameter of this distribution have precise ecological interpretations, the first being a measure of the excess of either rare or common species, and the second as a quantification of the proportion of persistent species in the focal community. Applying the Weibull model to 534 global tree communities we demonstrate that plots of the parameters of the Weibull distribution demark ecologically impossible species abundance distributions. This promises new insight into the ecological constraints on community assembly.

**Keywords:** Preston plot, species abundance distribution, species assembly, statistical fitting, Weibull distribution, Whittaker plot

## 1. Introduction

### 1.1 Theoretical Background

Within ecological assemblages, species generally differ widely in abundances (Magurran, 2004; Matthews and Whittaker, 2015). Often, the dominant species exceed the least abundant species by more than five orders of magnitude (Ulrich et al., 2010). Since Motomura (1932) formally introduced the concept of the species relative abundance distribution (SAD), the question of which ecological processes are responsible for this large variance in abundance has been controversially discussed (Tokeshi, 1999; Hubbell, 2001; McGill et al. 2007; Locey and White, 2013). Whilst early SAD models focused on the role of niche (Sugihara, 1980) vs. stochastic processes (May, 1975), recent discussion has largely centered on the impact on the SAD of dispersal (Hubbell, 2001; Dexter et al., 2017) vs. persistence (Magurran and Henderson, 2003). These discussions have sparked the development of a large number of different SAD models, each based on a specific set of assumptions about community assembly (Magurran, 2005; McGill et al., 2007). Irrespective of the ecological and theoretical background, SADs are commonly used for biodiversity assessment and monitoring (Matthews and Whittaker, 2015).

Models of relative abundances need to be fitted to observed distributions, a task that is far

from being straightforward (Ulrich et al., 2010; Mathews and Whittaker, 2014; Baldrige et al., 2016). To be of ecological value these fits need to be compared among different assemblages, for instance to infer environmental or geographical gradients (Ulrich et al., 2016a, b) and variation in abiotic conditions and biotic interactions (Erlén and Morris, 2015). Traditionally, SAD models have mostly been fitted to the statistical distribution (Matthews and Whittaker, 2014), where abundances are classified into  $\log_2$  bins prior to fitting (Preston, 1948, Fig. 1 left panels). As the selection of the logarithmic base and the way in which abundances are binned are arbitrary decisions, different binning procedures might have major impacts on model fitting (Nekola et al., 2008; Connolly and Dornelas, 2011). Further, any sufficient estimate of class frequencies needs an appropriate number of species, making frequency distribution fits reliable only for larger communities (Wilson et al., 1993). In contrast, plots that use ranked abundances for all species (Whittaker plots: Whittaker, 1975; Bazzaz, 1975) do not lose information due to data binning (Fig. 1 right panels) or the pooling of species to observed numbers of individuals. As a consequence, fits become reliable at much lower species richness. Ulrich et al. (2010) recommended ten species as the lower boundary for model fit, whereas Wilson (1993) reported that even 40 species may not be enough to reliably identify particular types of statistical distributions. Importantly, deviation of the fits from the original rank – abundance plot can be directly traced down to single species and possibly to the underlying processes that influence the abundance of these species. However, rank – abundance fits might fail if an excess of few very abundant or very rare species biases the metric used for fitting (often ordinary least squares in combination with Akaike information maximization). Further, the octaves of the SAD models that are based on a statistical distribution (e.g. the lognormal distribution) and not on an algorithm that directly generates for each species the expected abundance (e.g. most niche division based models, Tokeshi, 1996) need to be interpolated to species abundances prior to fitting, introducing a degree of subjectivism.

As observed SADs differ widely in shape and scale (the range in abundances), current models, that are based on one (shape, for instance the lognormal and most niche division models) or two (shape and scale, particularly the log-series) parameters, have limited variation in shape. Most are not able to accurately mimic all of the main different SAD types and do not cover the whole possible SAD space (i.e. all of the different empirically observed SAD forms). For instance, in recent global comparisons of dryland plant (Ulrich et al., 2016a) and forest tree (Ulrich et al., 2016b) communities, the latitudinal variation in SAD shape required separate fits of two different models. In this situation it became difficult to identify

any gradual latitudinal trend. In this respect, Baldrige et al. (2016) reported common distribution based SAD models had weak discrimination power, making any ecological inference challenging.

These problems regarding model fitting and comparison, the biases introduced by the binning of distributions, and the need for multiple model fits call for the development of flexible statistical SAD descriptors that are able to mimic various SAD shapes. In this respect, Ulrich et al. (2010) demonstrated that SADs can generally be classified into three basic shapes. The first shape is a lognormal type statistical distribution characterized by a larger number of species with intermediate abundance and fewer species with high and low abundance (Fig. 1a). In empirical SADs, there is often an excess of rare species than predicted by a lognormal distribution (Fig. 1b). The second shape is equivalent to a log-series sample distribution (Fisher et al., 1943), which is characterized by a few abundant and a larger number of relatively rare species (Fig. 1c). A small number of assemblages, particularly arthropod samples, follow a third shape, a power function SAD characterized by a heavy tail of rare species (Pueyo, 2006; Ulrich et al., 2010) (Fig. 1d). Here, we argue that a versatile SAD model must be able to fit these three basic shapes.

Hughes (1986) was the first to develop a flexible SAD model. However, fitting this model is not straightforward and the model has received little attention. Tokeshi (1996) described a one parameter resource division model that is able to fit the symmetric and skewed lognormal, but not other SAD shapes (Fig 1). The dynamic model of Dewdney (2000) provides good fits to log-series shaped distributions and possibly also to power functions. Ugland et al. (2007) developed the Gambin model, based on a discrete version of the statistical gamma distribution. These authors argued that Gambin is able to mimic several observed distribution shapes by variation of a single parameter ( $\alpha$ ); small values of  $\alpha$  characterize log-series SAD shapes, while higher values indicate lognormal curve shapes. As Gambin is intended to be a descriptor only, the parameter has no clear ecological interpretation, rather, it is simply a measure of the shape of the SAD. In addition, Gambin is based on a statistical distribution (as used by Preston 1948) and involves binning the abundance data into octaves before fitting. There is no straightforward way to rescale the Gambin distribution to species abundances.

Of course, neutral, ecological drift models (Hubbell, 2001) provide a mechanistic interpretation of observed abundance distributions with ecologically well-defined parameters. Depending on the probability of dispersal, speciation rates, local abundances, and meta-community size they provide predictions (Fig. 1) that are close to those of either the skewed

lognormal or log-series (Hubbell, 2001). These models do not predict power function SADs (Fig. 1), commonly observed in arthropods (Siemann, 1999; Borda-de-Água et al., 2017). In addition, neutral models are notoriously difficult to fit and certain models require information on the structure of the underlying meta-community (Rosindell et al., 2010). Maximum likelihood fitting of neutral models requires complex assumptions about macroevolutionary processes (Etienne et al., 2007) that can result in unfeasible (or unmeasurable) parameter values, such as speciation rates (Ricklefs, 2003) and long-distance dispersal (Rosindell and Cornell, 2009). These issues have limited the application of neutral models for the prediction of species abundances.

Here, we strongly argue that a flexible descriptive SAD model needs to predict the abundances of each species directly and that abundance – rank orders are superior to distribution approaches. Only such species-focused models make it possible to trace deviations from observation and possible ecological drivers directly to the species level. There is also clearly a need to link a SAD model to species functional traits or phylogenetic relationships, for instance to infer how traits influence the dominance order of species abundances (Jones et al., 2017). Consequently, a flexible SAD model should ideally be based on Whittaker plot data.

Stauffer (1979) was apparently the first to propose the Weibull distribution (Weibull, 1951) as a model to explain observed species abundance distributions in forest trees. However, his derivation of the model and parameter interpretation were based on the broken stick model (Mac Arthur, 1957), a model that was subsequently found to be unrealistic. As his approach was distribution based the work did not receive the attention it deserves. Recently, Storch et al. (2018) used the one-parameter Weibull distribution to mimic the log-series. The R package SADs (Prado et al. 2017) provides functionality to fit the Weibull model to the statistical distribution, but not to rank abundances.

Here, we reintroduce the Weibull distribution as a flexible descriptive model that meets the above defined requirements. We show that this distribution is able to fit the most commonly observed SAD shapes. This ability enables us to compare the respective parameter values across SADs from different studies within the same modelling framework. Although Weibull distributions result from several stochastic processes (Rinne, 2008), we do not claim that the model is directly linked to ecological processes. We also demonstrate that the shape and the scale parameters of the model have straightforward ecological interpretations that can be used in ecological analyses. As a case study, we illustrate our approach using a set of global forest tree data.

## 1.2 The Weibull distribution as a SAD model

The empirical Weibull distribution (Weibull, 1951; Rinne, 2008) is an extension of the exponential family of distributions and is widely used in survival analyses (Lawless, 2003) and extreme value forecasting (Carter and Challenor, 1983). Its two-parameter form has the probability density function (pdf)

$$p(x > 0; \eta; \lambda) = \frac{\eta}{\lambda} \left(\frac{x}{\lambda}\right)^{\eta-1} e^{-\left(\frac{x}{\lambda}\right)^{\eta}} \quad (1)$$

where  $\eta$  is the shape and  $\lambda$  the scale parameter. When applied to species abundances the random variate  $x$  must contain log-transformed values. Here, we use the  $\ln$ -transform as a standard.

Both parameters,  $\eta$  and  $\lambda$ , have clearly defined ecological interpretations. The scale parameter is given by

$$\lambda^2 = \sigma^2 \left[ \Gamma\left(1 + \frac{2}{\eta}\right) - \left(\Gamma\left(1 + \frac{1}{\eta}\right)\right)^2 \right]^{-1} \quad (2)$$

where  $\Gamma$  denotes the gamma function and  $\sigma^2$  the variance. Therefore,  $\lambda$  increases with increasing variance in abundance and provides a measure of the range in  $\ln$ -transformed species abundances. We note that the abundance range is closely connected to the concept of evenness. Evenness measures the variance in abundance (Smith and Wilson 1996). Therefore, the wider the range in abundance is, the lower is the degree of evenness. The  $\lambda$  parameter can therefore be interpreted as a measure of SAD shape specific evenness. The shape parameter  $\eta$  is connected to the excess of either highly abundant species (low  $\eta$ ) or rare species (high  $\eta$ ).

The Weibull distribution is most often applied to assess prospected survival and failure times in demography and industry. If  $x$  denotes the time to failure the quantity

$$T = \lambda \Gamma\left(\frac{1}{\eta} + 1\right) \quad (3)$$

is an estimate of the expected average time to failure (extinction in ecological terms). This interpretation can be extended to species abundances. Abundant species should be highly competitive or adapted to a focal habitat. Under this interpretation, the average value  $T$  divides those species that are more competitive (having higher reproductive output) from those that are less competitive (lower reproductive output). The position of  $T$  along the abundance axis in Preston plots (Fig. 1, left panels) indicates therefore the proportion of species that are relatively competitive. Interpreting the parameters of the Weibull distribution in terms of reproductive output and therefore fitness implies that high values of  $T$  indicate

increased proportions of species with high fitness.

Using Weibull distributed random numbers (Press et al., 1986) rescaled to relative abundances, our approach predicts directly the abundances of each species without the need for retransformation. Further, Tokeshi (1999) advised the use of such stochastic models for which goodness of fit can be assessed in terms of standard errors for the predicted abundance of each species. As our model involves randomly assigned abundances, we here obtain such errors from 100 fits to each empirical SAD and assess goodness of fit from the proportion of species falling outside the 95% confidence limits of the model.

## 2. Methods

### 2.1 Fitting the Weibull model to empirical species abundance distributions

Simple but effective maximum likelihood estimators for the two-parameter Weibull distributions already exist (Nwobi and Ugomma, 2014). However, predicting the abundances of each species requires interpolation of the distribution, making the resultant fits less reliable. Here we use a twofold approach to fitting the Weibull distribution to SAD data that directly generates expected species abundances without the need of back-binning the distribution data. We iteratively encapsulate the parameter values of  $\eta$  to find the value that minimizes a goodness of fit metric defined by the reduced major axis value

$$fit = \frac{\sum_{i=1}^S (lnA_{i,obs} - lnA_{i,pred})^2 + \min[(j-i)^2]}{S} \quad (4)$$

where  $A_{i,obs}$  and  $A_{i,pred}$  are the respective observed and predicted relative abundances of species  $i$  in the community of  $S$  species, and  $j$  runs over all  $S$  species. At each step, this fitting process involves estimating the expected abundances  $A_{i,pred}$  obtained from Weibull distributed random numbers ( $f_W$ , see Press et al., 1986)

$$f_{W;k} = \left( \frac{-1}{\lambda^{-\eta_{k-1}}} \ln(1 - f_l) \right)^{1/\eta_{k-1}} \quad (5)$$

where  $f_l$  is a linear random number between 0 and 1. Eq. 5 uses an initial maximum likelihood estimate of  $\lambda$ , which is given by

$$\lambda = \left( \frac{1}{S} \sum_{i=1}^S x_i^\eta \right)^{1/\eta} \quad (6)$$

where  $\eta$  comes from the solution of

$$\frac{1}{S} \frac{\sum_{i=1}^S x_i^\eta \ln x_i}{\sum_{i=1}^S x_i^\eta} - \frac{1}{\eta} - \frac{\sum_{i=1}^S \ln x_i}{S} = 0 \quad (7)$$

$S$  is the observed number of species (Cohen, 1965). Eq. 7 can easily be solved using common numerical methods. The values



$$A_i = e^{x_i} \quad (8)$$

are then the Weibull abundance estimates of each species  $i$ .

In contrast to Ulrich et al. (2016a, b), here we have used reduced major axis fits rather than OLS fits as the former places similar weight on all species irrespective of relative abundance, whilst the latter approach places greater weight on abundant and rare species (Connolly and Dornelas, 2011). We note that OLS fits to the same data returned qualitatively identical results.  $fit < 0.05$  indicates an excellent fit while  $fit > 0.3$  is poor. The Fortran code used for fitting has already been published in Ulrich et al. (2016a, b) and is freely available from WU on request.

## 2.2 Case study

To illustrate our fitting approach, we fitted the Weibull distribution to four artificial communities of 50 species each that are presented in Figure 1. These represent the major SAD shapes observed in nature. In a case study using empirical data, we fitted the Weibull distribution to a set of 534 fully censused tree communities. These datasets were compiled by restricting the global compilation of tree communities published in Ulrich et al. (2016) and Kubota et al. (2018) to 534 fully censused communities ranging between 10 and 100 species and containing between 20 and 38902 individuals. These communities span a wide range of different abundance distributions (Ulrich et al. 2016). Evenness  $J$  of these communities was calculated from  $J = H/\ln(S)$ , where  $H$  denotes the Shannon diversity.

## 3. Results

### 3.1 Performance of the Weibull model

The Weibull model provided a nearly perfect fit ( $fit = 0.004$ ) to the symmetrical lognormal data (Fig. 2a) and very good approximations to the power function data ( $fit = 0.007$ ) (Fig. 2d). The fits to the left skewed lognormal data with an excess of rare species ( $fit = 0.05$ ) (Fig. 2b) and the log-series data ( $fit = 0.03$ ) (Fig. 2c) were only marginally weaker. Importantly, using the parameters of the Weibull model it was clearly possible to separate these four common types of SAD (Table 1). For the lognormal (Fig. 2a), the log-series (Fig. 2c) and the power function data (Fig. 2c) all of the observed species abundances in Fig. 2 A (left panels) were within the 95% confidence limits of the model fits, whilst for the skewed lognormal data (Fig. 2b) three of the 50 species (6.0%) deviated from the fitted value. The rescaling of the distributions and the fits to  $\log_2$  frequency distributions (Fig. 2 B) confirmed the good performance of the Weibull distribution.

The 95% bootstrap confidence limits of the basic parameters of the fitted distribution (skewness, kurtosis, coefficients of variation in log-abundance), as well as the Weibull model parameters  $\lambda$  and  $\eta$  included in all but three cases the observed values (Tab. 1). In addition, the parameters significantly differed (one-way ANOVA,  $P < 0.001$ ) between the four shapes (Tab. 1), demonstrating the excellent discriminatory power of the model.

### 3.2 Abundance distributions of global tree communities

Ulrich et al. (2018) provide fits to all 534 distributions in a figshare database (10.6084/m9.figshare.5975098) demonstrating the excellent performance of the model. Weibull fits to 144 of the 534 communities (27.0%) were excellent ( $fit < 0.05$ , Fig. 3), while only 20 fits (3.7%) were comparatively poor ( $fit > 0.3$ , Fig. 3 and Ulrich et al. 2018). Goodness of fit was independent of species richness (Fig. 3a) but moderately increased with increasing total abundance (Fig. 3b), abundance range (Fig. 3c) and model parameters (Figs 3 d, e). Goodness of fit decreased with increasing skewness, that is, the excess of abundant species (Fig. 3f).

These results are corroborated by the fact that observed skewness and modelled skewness were strongly positively correlated (Fig. 4a). As expected from eq. 2, the scale parameter  $\lambda$  was positively correlated to the observed abundance range (Fig. 4b). Low  $\eta$  values were linked to an excess of rare species, and high  $\eta$  values to an excess of abundant species (Fig. 4c). Values of  $\lambda$  and  $\eta$  were linearly correlated (Fig. 4d), defining areas of not realized parameter combinations. Although  $\lambda$  and  $\eta$  are related to the variance and skewness of the SAD, respective skewness - variance plots performed worse in identifying realized and forbidden SAD shapes (Fig. 5e).

## 4. Discussion

Our fits to 534 global tree communities showed that the two-parameter Weibull distribution is an appropriate tool to mimic a wide variety of observed species abundance distributions (Ulrich et al. 2018). We note that the model does not provide a mechanistic explanation for these distributions although several stochastic processes are known to be Weibull distributed (Rinne, 2008). The data presented in Ulrich et al. (2018) also show that the model might provide weaker fits in communities dominated by a small number of highly abundant species and in communities with a marked excess in very rare species. We note that this fitting problem also applies to other SAD models designed to fit specific stochastic niche

partitioning processes (Tokeshi 1999) and statistical distributions (Ulrich et al. 2010).

An important finding of the present study regards the relationship of the shape and the scale parameters of the Weibull distribution that for the first time enable us to define limits to the observed shapes of empirical abundance distributions (Fig. 4d). Respective plots of skewness and variance (Fig. 4e) and skewness and evenness (Fig. 4f) did not recover these limits. With few exceptions, communities with  $\eta > 3$  did not exist. As  $\lambda$  increases with abundance range (Fig. 4b), communities with  $\lambda > 6$  might exist for instance in invertebrate assemblies with abundance ranges  $>$  six orders of magnitude. High  $\eta$  is linked to the excess of relatively rare species (Tab. 1). The theoretical power fraction SAD with a high excess of such species (Figs. 1 and 2) had  $\eta > 3$ . Therefore, our result strongly indicates that  $\eta$  cannot exceed a certain limit in natural communities and that such extreme SADs do not exist in global forest tree communities.

What do such ‘forbidden’ communities look like? Fig. 5 shows that low  $\eta$  in combination with low  $\lambda$  (Fig. 5a) predicts communities with an initial steep decline in abundance and a short ‘heavy tail’ of rare species. High  $\eta$  and low  $\lambda$  (Fig. 5b) generate SADs very similar to the popular, but discredited, broken stick distribution (MacArthur, 1957). Our results question whether such communities are realized. Indeed, empirical evidence for broken stick communities is very limited (Smart, 1976). Higher values of  $\lambda$  and  $\eta$  generate communities for which the proportion of very rare species exceeds 40% (Fig. 5c). As already noted, such distributions do not exist for species rich communities.

The fact that  $\eta$  was bounded to values between 1 and 3 implied that  $\Gamma(1/\eta+1)$  takes values between 0.89 and 1, making the failure time  $T$  to be a nearly linear function of  $\lambda$  (eq. 3). This fact gives the scale parameter  $\lambda$  yet another interpretation. It denotes the relative proportion of adapted and possibly persistent species in a community. In the forest data,  $T$  was significantly linearly correlated with the total abundance of all species ( $r^2 = 0.54$ ) and abundance range ( $r^2 = 0.58$ ), but not with species richness ( $r^2 = 0.01$ ). We speculate that communities with high overall abundance also contain a higher proportion of species not endangered by local extinction.

The approximately linear relationship between  $\eta$  and  $\lambda$  makes the respective quotient  $q = \eta/\lambda$  a potentially new metric of community composition that catches the relationship between range in abundance (Fig. 4b) and the excess of either rare or abundant species (Fig. 4c). It is beyond the scope of the present paper to apply this metric to environmental and trait data but we note that in the forest tree communities of this study,  $q$  was positively correlated to

community evenness ( $r^2 = 0.37$ ) and consequently negatively to the range of abundances ( $r^2 = 0.71$ ). This opens the possibility of partitioning dissimilarity in abundances into two parts, one part linked to SAD shape (skewness) and a second part linked to community evenness.

Finally, the good fits of the Weibull distribution to observed SADs (Fig. 3 and Ulrich et al. 2018) makes it possible to take a species based approach to SADs to compare model prediction and observation. In this respect, distributions of functional traits and phylogenetic relationships that characterize the relative abundance of realized species characteristics in a community have come into focus (Gross et al. 2017; Jones et al., 2017). Being explicitly based on species – rank order abundances, they allow for a direct comparison of species traits among communities and between realized and theoretically expected values (Cornwall and Ackerly, 2009). As neutral model fitting is challenging in most cases (Gotelli and McGill, 2006; Rosindell et al., 2010), the application of the Weibull distribution offers a valuable alternative, particularly for the identification of outliers in abundance and the analysis of how such outlying species influence community trait space and functional performance. In this respect, Ulrich et al. (2016a) and Matthews et al. (2017) have recently shown that comparative analyses of SAD shape and parameters can reveal important biogeographic patterns (e.g. latitudinal and climate gradients). Our approach not only provides a flexible model but offers ecologically interpretable parameters that can be related to species trait and environmental data. We argue that the Weibull distribution might form a statistical standard (similar to a null model) to which observed species relative abundances can be compared.

## Acknowledgements

W.U. was supported by the Polish National Science Centre (grant 2014/13/B/NZ8/04681). Y.K. was support by the Japan Society for the Promotion of Science (no. 15K14607). R.N. acknowledges support by the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers of the Japan Society for the Promotion of Science.

## Data accessibility

Raw data of the forest plots used in this study are contained in Ulrich et al. (2018).

## References

1. Baldridge, E., Harris, D.J., Xiao, X., White, E.P., 2016. An extensive comparison of

- species-abundance distribution models. *PeerJ* 4, e2823.
2. Bazzaz F.A., 1975. Plant species diversity in old-field successional ecosystems in southern Illinois. *Ecology* 56, 485–488.
  3. Borda-de-Água, L., Whittaker, R.J., Cardoso, P., Rigal, F., Santos, A.M.C., Amorim, I.R., Parmakelis, A., Triantis, K.A., Pereira, H.M., Borges, P.A.V., 2017. Dispersal ability determines the scaling properties of species abundance distributions: a case study using arthropods from the Azores. *Sci. Rep.* 7, 3899.
  4. Carter, D.J.T., Challenor, P.G., 1983. Application of extreme value analysis to Weibull data. *Quart. J. R. Meteorolog. Soc* 109, 429-433.
  5. Cohen, A.C., 1965. Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics* 7, 3.
  6. Connolly, S.R., Dornelas, M., 2011. Fitting and empirical evaluation of models for species abundance distributions. In: Magurran, A.E. and McGill, B.J. (eds.) *Biological diversity: frontiers in measurement and assessment*, Oxford University Press, pp. 123–140.
  7. Cornwall, W., Ackerly, D., 2009. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecol. Monogr.* 79, 109-126.
  8. Dewdney, A.K., 2000. A dynamical model of communities and a new species–abundance distribution. *Biol. Bull.* 198, 152–165.
  9. Dexter, K.G., Lavin, M., Torke, B.M., Twyford, A.D.D., Kursar, T.A., Coley, T.D., Drake, C., Hollands, R., Pennington, R.T., 2017. Dispersal assembly of rain forest tree communities across the Amazon basin. *Proc. Natl. Acad. Sci. USA* 114, 2645-2650.
  10. Erlén, J., Morris, F.W., 2015. Predicting changes in the distribution and abundance of species under environmental change. *Ecol. Lett.* 18, 303-314.
  11. Etienne, R.S., Apol, M.E.F., Olff H., Weissing, F.J., 2007. Modes of speciation and the neutral theory of biodiversity. *Oikos* 116, 241-258.
  12. Fisher, R.A., Corbet, A.S., Williams, C.B., 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Animal Ecol.* 12, 42–58.
  13. Hubbell, S.P., 2001. *The unified theory of biogeography and biodiversity*. Princeton University Press.
  14. Hughes, R.G., 1986. Theories and models of species abundance. *Am. Nat.* 128, 879–899.

15. Gotelli, N.J., McGill, B.J., 2006. Null versus neutral models: what's the difference?  
Ecography 29, 793-800.
16. Gross, N., Le Bagousse-Pinguet, Y., Liancourt, P., Berdugo, M., Gotelli, N.J., Maestre,  
F.T., 2017. Functional trait diversity maximizes ecosystem multifunctionality. Nature  
Ecol. Evol. 1, 0132.
17. Jones, S.K., Ripplinger, J., Collins, S.L., 2017. Species reordering, not changes in  
richness, drives long-term dynamics in grassland communities. Ecol. Lett. 20, 1556-  
1575.
18. Kubota Y., Kusumoto B., Shiono T., Ulrich W., 2018. Multiple filters affect tree  
species assembly in mid-latitude forest communities. Oecologia, in press.
19. Lawless, J. F., 2003. Statistical Models and Methods for Life time Data. 3<sup>rd</sup> Edition,  
John Wiley and Sons, New York.
20. Locey, K.J., White, E.P., 2013. How species richness and total abundance constrain the  
distribution of abundance. Ecol. Lett. 16, 1177–1185.
21. MacArthur, R.H., 1957. On the relative abundance of bird species. Proc. Natl. Acad.  
Sci. USA 43, 293–295.
22. Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Science, Oxford.
23. Magurran, A., 2005. Species abundance distributions: pattern or process? Funct. Ecol.  
19, 177-181.
24. Magurran, A.E., Henderson, P.A., 2003. Explaining the excess of rare species in  
natural species abundance distributions. Nature 422, 714–716.
25. Matthews, T.J., Whittaker, R.J., 2014. Fitting and comparing competing models of the  
species abundance distribution: assessment and prospect. Front. Biogeogr. 6, 67–82.
26. Matthews, T.J., Whittaker, R.J., 2015. On the species abundance distribution in  
applied ecology and biodiversity management. J. Appl. Ecol. 52, 443–454.
27. Matthews, T.J., Borges, P.A.V., de Azevedo, E.B., Whittaker, R.J., 2017. A  
biogeographical perspective on species abundance distributions: recent advances and  
opportunities for future research. J. Biogeogr. 44, 1705–1710
28. May, R.M., 1975. Patterns of species abundance and diversity. In: Cody, M.L.,  
Diamond, J.M. (eds.) Ecology and Evolution of Communities, Cambridge University  
Press, pp. 81–120.
29. McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K.,  
Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E.,  
Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P.,

2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–1015.
31. Motomura, I., 1932. A statistical treatment of associations. *Japan. J. Zool.* 44, 379–383 (in Japanese).
32. Nekola, J.C., Šizling, A.L., Boyer, A.G., Storch, D., 2008. Artifacts in the log-transformation of species abundance distributions. *Folia Geobot.* 43, 259–268.
33. Nwobi, F.N., Ugomma, C.A., 2014. A Comparison of Methods for the Estimation of Weibull Distribution Parameters. *Metodologiya Zvezki* 11, 65–78.
34. Prado, P.I., Dantas Miranda, M., Chalom, A. 2017. Package sads. <http://piLaboratory.github.io/sads>.
35. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1986. Numerical Recipes, The Art of Scientific Computing, Cambridge University Press.
36. Preston, F.W., 1948. The commonness, and rarity, of species. *Ecology* 29, 254–283.
37. Pueyo, S., 2006. Self-similarity in species–area relationship and in species abundance distribution. *Oikos* 112, 156–162.
38. Ricklefs, R. E., 2003. A comment on Hubbell’s zero-sum ecological drift model. *Oikos* 100, 185–192.
39. Rinne, H., 2008. The Weibull Distribution. A Handbook. CRC Press.
40. Rosindell, J., Cornell, S.J., 2009. Species–area curves, neutral models, and long-distance dispersal. *Ecology* 90, 1743–1750.
41. Rosindell, J., Cornell, S.J., Hubbell, S.P., Etienne, R.S., 2010. Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.* 13, 716–727.
42. Smart, J.S., 1976. Statistical tests of the broken-stick model of species-abundance relations. *J. Theor. Biol.* 59, 127–139.
43. Siemann, E., Tilman, D., Haarstad, J., 1999. Abundance, diversity and body size: Patterns from a grassland arthropod community. *J. Animal Ecol.* 68, 824–835.
44. Smith, B, Wilson, J.B., 1996. A consumer’s guide to evenness indices. *Oikos* 76, 70–82.
45. Stauffer, H.B., 1979. A derivation for the Weibull distribution. *J. Theor. Biol.* 81, 55–63.
46. Storch, D., Bohalková, E., Okie, J., 2018. The more-individuals hypothesis revisited: the role of community abundance in species richness regulation and the productivity–diversity relationship. *Ecol. Lett.* 21, 920–937.
47. Tokeshi, M., 1996. Power fraction: a new explanation for species abundance patterns

in species-rich assemblages. *Oikos* 75, 543–550.

49. Tokeshi, M., 1999. Species coexistence: ecological and evolutionary perspectives. Blackwell Scientific, Oxford.

50. Ugland, K.I., Lambshead, P.J.D., McGill, B., Gray, J.S., O’Dea, N., Ladle, R.J., Whittaker, R.J., 2007. Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evol. Ecol. Res.* 9, 313-324.

51. Ulrich W., Olrik M., Ugland, K.I., 2010. A meta-analysis of species - abundance distributions. *Oikos* 119, 1149-1155.

52. Ulrich, W., Kusumoto, B., Shiono, T., Kubota, Y., 2016a. Climatic and geographical correlates of global forest tree species abundance distributions and community evenness. *J. Veg. Sci.* 27, 295-305.

53. Ulrich, W., Soliveres, S., Thomas, A.D., Dougill, A.J., Maestre, F.T., 2016b. Environmental correlates of species rank – abundance distributions in global drylands. *Persp. Plant Ecol. Evol. Syst.* 20, 56-64.

54. Ulrich, W., Nakadai, R., Matthews, T., Kubota, Y. 2018. Weibull fits to global tree species abundance distributions. figshare [https://figshare.com/articles/Weibull\\_fits/5975098](https://figshare.com/articles/Weibull_fits/5975098).

55. Weibull, W., 1951. A Statistical Distribution of wide Applicability. *J. Appl. Mech.* 18, 239–296.

56. Whittaker, R.H., 1975. *Communities and Ecosystems*, 2nd edn. MacMillan, New York.

57. Wilson, J.B., 1993. Would we recognise a broken-stick community if we found one? *Oikos* 67, 181-183.

## **Software**

Fortran source code and the updated stand-alone application RAD 2.0 (Ulrich et al. 2010, 2016a, b) are freely available from W.U. by request.

## **Data availability**

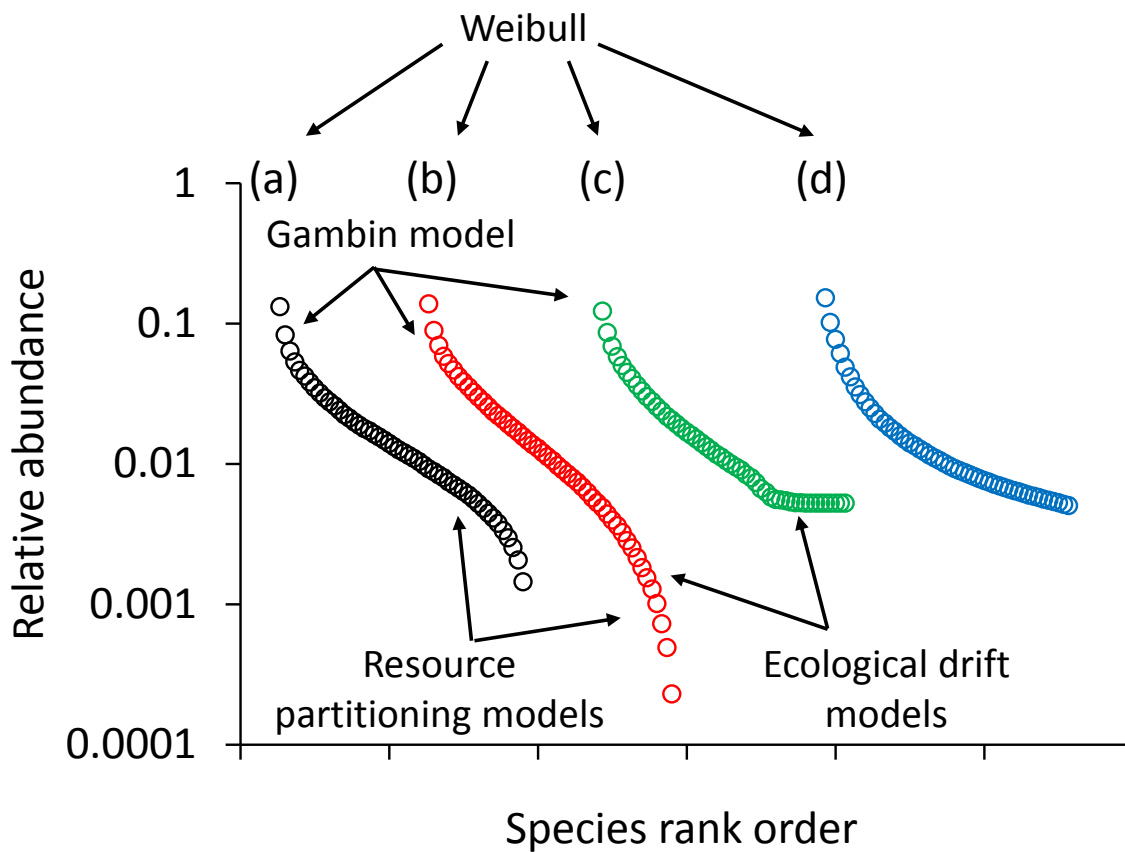
Weibull fits to global tree species abundance distributions. figshare [https://figshare.com/articles/Weibull\\_fits/5975098](https://figshare.com/articles/Weibull_fits/5975098).



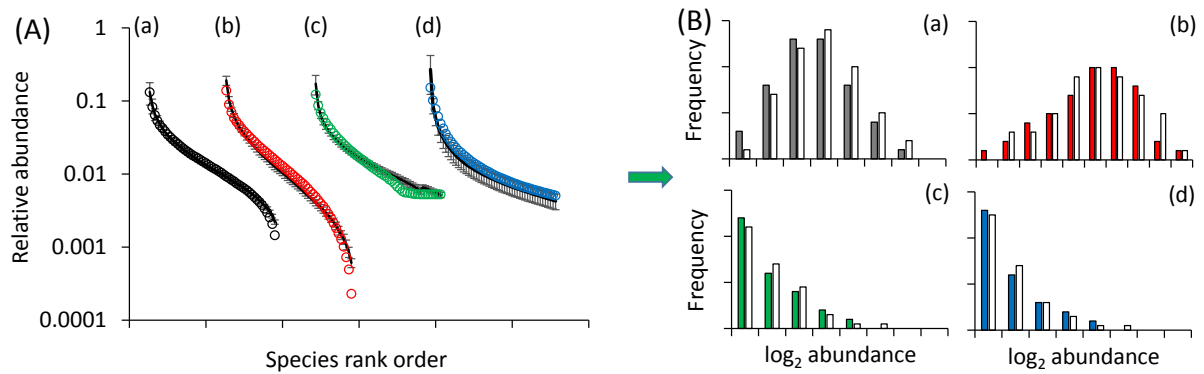
**Table 1.** Coefficients of variation (CV) and moments of the ln-transformed relative abundances of the communities in Fig. 2, and the respective fits of the Weibull distribution (given are shape and scale parameters of the best fit) together with upper and lower two-sided 95% confidence limits (CL) of 1000 bootstrap samples. Shape and scale of the data refer to maximum likelihood Weibull parameters taken directly from the ln-transformed abundances.

| Model            | Statistics | CV          | Standard deviation | Skewness     | Kurtosis | Shape | Scale |
|------------------|------------|-------------|--------------------|--------------|----------|-------|-------|
| lognormal        | Data       | 0.45        | 1.41               | 0.10         | -0.21    | 2.51  | 2.45  |
|                  | fit        | 0.53        | 1.35               | 0.36         | -0.26    | 2.45  | 2.06  |
|                  | lower CL   | 0.43        | 1.12               | -0.10        | -1.32    | -     | -     |
|                  | upper CL   | 0.64        | 1.61               | 0.87         | 0.50     | -     | -     |
| skewed lognormal | Data       | <b>0.38</b> | 2.00               | <b>-0.50</b> | 0.05     | 3.22  | 4.28  |
|                  | fit        | <b>0.54</b> | 1.87               | <b>0.12</b>  | -0.39    | 3.16  | 3.23  |
|                  | lower CL   | <b>0.43</b> | 1.58               | <b>-0.35</b> | -1.25    | -     | -     |
|                  | upper CL   | <b>0.65</b> | 2.23               | <b>0.62</b>  | 0.25     | -     | -     |
| log-series       | Data       | <b>0.98</b> | 1.27               | 0.75         | -0.39    | 1.11  | 0.93  |
|                  | fit        | <b>0.80</b> | 1.14               | 1.03         | 0.90     | 1.09  | 0.94  |
|                  | lower CL   | <b>0.65</b> | 0.91               | 0.47         | -1.35    | -     | -     |
|                  | upper CL   | <b>0.96</b> | 1.42               | 1.72         | 2.78     | -     | -     |
| power function   | Data       | 0.91        | 1.21               | 1.20         | 0.89     | 1.15  | 0.96  |
|                  | fit        | 0.92        | 1.23               | 1.66         | 3.40     | 1.12  | 0.96  |
|                  | lower CL   | 0.73        | 0.87               | 0.93         | -0.71    | -     | -     |
|                  | upper CL   | 1.14        | 1.62               | 2.66         | 7.33     | -     | -     |

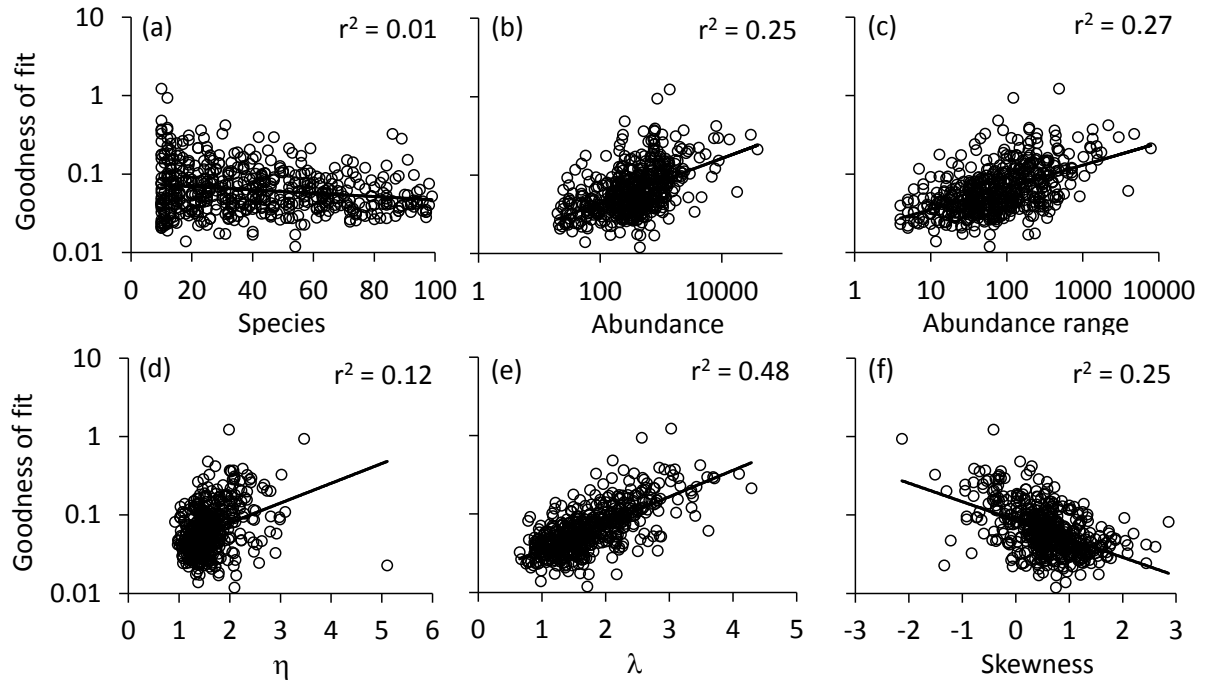
**Figure 1.** Four typical shapes of species abundance distributions (Whittaker species - rank order plots). a) a lognormal distribution, b) left skewed lognormal with an excess of rare species, c) a log-series distribution, d) a power function. Each community contains 50 species. Arrows show which types of distributions the selected flexible SAD models can mimic.



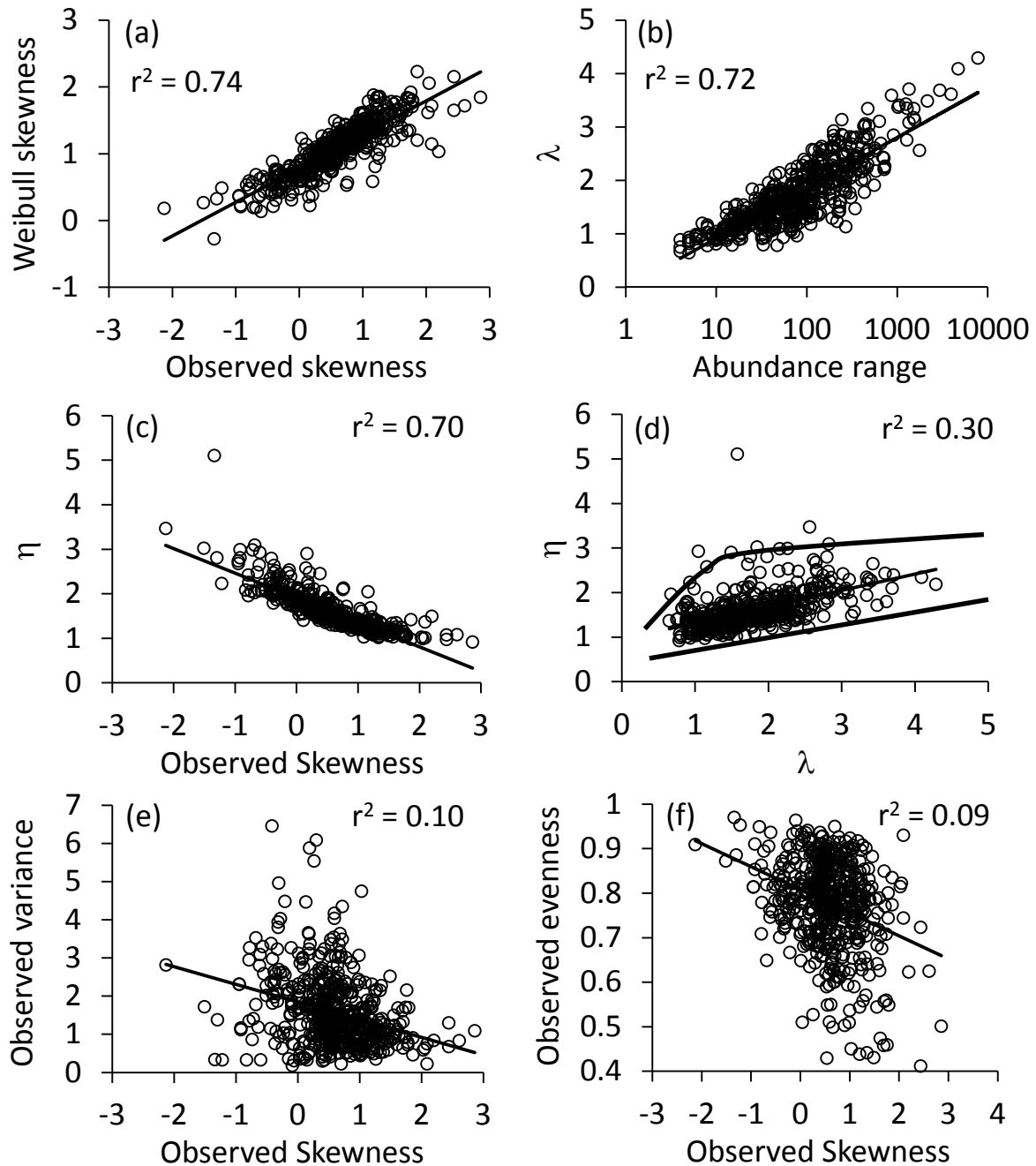
**Figure 2.** Reduced major axis fits of the Weibull model to rank abundance (left side) and the respective  $\log_2$ -binned distribution data (right side) of the four typical shapes of species rank order –abundance distributions, of 50 species each shown in Fig. 1. a) a lognormal distribution, b) left skewed lognormal with an excess of rare species, c) a log-series distribution, and d) a power function. Error bars denote one standard deviation of the Weibull model fits. Parameter values are given in Tab. 1



**Figure 3.** Goodness of fit of 534 empirical global forest tree communities in relation to species richness (a), total abundance (b), range in abundance (c), Weibull fit parameters  $\lambda$  (d), and  $\eta$  (e), and observed skewness (f). Given are the  $r^2$  values from exponential (a, d, e, f) and power function (b, c) OLS regressions.



**Figure 4.** Relationships between observed and expected skewness (a),  $\lambda$  and observed abundance range (b),  $\eta$  and observed skewness (c),  $\eta$  and  $\lambda$  (d), observed skewness and variance (e), and evenness (f) from the fits of the Weibull distribution to 534 empirical global tree communities. Given are the  $r^2$  values of linear (a, c, d, e) and logarithmic (b) OLS regressions. Approximate upper and lower boundaries of  $\eta$  are given in (d).



**Figure 5.** Three examples of communities (50 species each) generated by Weibull distributions that are not realized according to Figure 4d.

